

UIT-R

Secteur des Radiocommunications de l'UIT

Recommandation UIT-R BT.1908-0 (01/2012)

Techniques de mesure objective de la qualité vidéo pour les applications de radiodiffusion utilisant la télévision haute définition en présence d'un signal de référence réduit

Série BT
Service de radiodiffusion télévisuelle



Union
internationale des
télécommunications

Avant-propos

Le rôle du Secteur des radiocommunications est d'assurer l'utilisation rationnelle, équitable, efficace et économique du spectre radioélectrique par tous les services de radiocommunication, y compris les services par satellite, et de procéder à des études pour toutes les gammes de fréquences, à partir desquelles les Recommandations seront élaborées et adoptées.

Les fonctions réglementaires et politiques du Secteur des radiocommunications sont remplies par les Conférences mondiales et régionales des radiocommunications et par les Assemblées des radiocommunications assistées par les Commissions d'études.

Politique en matière de droits de propriété intellectuelle (IPR)

La politique de l'UIT-R en matière de droits de propriété intellectuelle est décrite dans la «Politique commune de l'UIT-T, l'UIT-R, l'ISO et la CEI en matière de brevets», dont il est question dans la Résolution UIT-R 1. Les formulaires que les titulaires de brevets doivent utiliser pour soumettre les déclarations de brevet et d'octroi de licence sont accessibles à l'adresse <http://www.itu.int/ITU-R/go/patents/fr>, où l'on trouvera également les Lignes directrices pour la mise en oeuvre de la politique commune en matière de brevets de l'UIT-T, l'UIT-R, l'ISO et la CEI et la base de données en matière de brevets de l'UIT-R.

Séries des Recommandations UIT-R

(Egalement disponible en ligne: <http://www.itu.int/publ/R-REC/fr>)

Séries	Titre
BO	Diffusion par satellite
BR	Enregistrement pour la production, l'archivage et la diffusion; films pour la télévision
BS	Service de radiodiffusion sonore
BT	Service de radiodiffusion télévisuelle
F	Service fixe
M	Services mobile, de radiorepérage et d'amateur y compris les services par satellite associés
P	Propagation des ondes radioélectriques
RA	Radio astronomie
RS	Systèmes de télédétection
S	Service fixe par satellite
SA	Applications spatiales et météorologie
SF	Partage des fréquences et coordination entre les systèmes du service fixe par satellite et du service fixe
SM	Gestion du spectre
SNG	Reportage d'actualités par satellite
TF	Emissions de fréquences étalon et de signaux horaires
V	Vocabulaire et sujets associés

Note: Cette Recommandation UIT-R a été approuvée en anglais aux termes de la procédure détaillée dans la Résolution UIT-R 1.

Publication électronique
Genève, 2020

© UIT 2020

Tous droits réservés. Aucune partie de cette publication ne peut être reproduite, par quelque procédé que ce soit, sans l'accord écrit préalable de l'UIT.

RECOMMANDATION UIT-R BT.1908-0*

Techniques de mesure objective de la qualité vidéo pour les applications de radiodiffusion utilisant la télévision haute définition en présence d'un signal de référence réduit

(2012)

Domaine d'application

La présente Recommandation définit les méthodes à appliquer pour évaluer la qualité vidéo perçue des applications de radiodiffusion utilisant la télévision haute définition lorsqu'un signal de référence réduit est disponible.

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) qu'il n'est plus à démontrer depuis longtemps que la possibilité de mesurer automatiquement la qualité d'une séquence vidéo diffusée constitue un atout précieux pour le secteur d'activité concerné;
- b) que la Recommandation UIT-R BT.1683 décrit des techniques de mesure objective de la qualité vidéo perçue pour la télédiffusion numérique à définition normale en présence d'une image de référence réduite;
- c) que la Recommandation UIT-R BT.709 donne la valeur de paramètres des normes de télévision haute définition pour la production et l'échange international de programmes et la Recommandation UIT-R BT.500 décrit des méthodes d'évaluation subjective de la qualité d'image y compris la télévision haute définition;
- d) que la télévision haute définition (TVHD) est de plus en plus utilisée en radiodiffusion;
- e) que, sur la base des résultats figurant dans le rapport sur la TVHD envoyé par le Groupe VQEG, la Commission d'études 9 de l'UIT-T a élaboré la Recommandation UIT-T J.342, relative à la mesure objective de la qualité vidéo pour la TVHD en présence d'une référence réduite;
- f) qu'une mesure objective de la qualité vidéo perçue pour la TVHD peut compléter les méthodes d'évaluation subjective,

recommande

- 1** d'utiliser le modèle décrit dans l'Annexe 1 en ce qui concerne la mesure objective de la qualité vidéo perçue pour les applications de radiodiffusion utilisant la TVHD en présence d'un signal de référence réduit.

* La Commission d'études 6 des radiocommunications a apporté des modifications rédactionnelles à la présente Recommandation en février 2020 conformément aux dispositions de la Résolution UIT-R 1.

Annexe 1

1 Introduction

La présente Recommandation décrit une méthode de mesure de la qualité vidéo à utiliser pour les applications non interactives de télévision haute définition (TVHD) lorsqu'il est possible d'utiliser une méthode de mesure avec référence réduite. Le modèle a fait l'objet d'une comparaison avec les notes subjectives de qualité obtenues au moyen de la Recommandation UIT-R BT.500. Les analyses ont montré que la précision de ce modèle était équivalente à celle du PSNR.

Pour fonctionner correctement, le modèle avec référence réduite doit disposer de la séquence vidéo source non dégradée afin d'extraire des paramètres. Ces paramètres extraits ainsi que la séquence vidéo dégradée constituent les entrées du modèle avec référence réduite. La méthode d'évaluation comporte un étalonnage (réglage du gain/décalage et alignement spatial/temporel) et une évaluation objective de la qualité vidéo.

Les séquences utilisées pour les tests de validation contenaient à la fois des dégradations dues au codage UIT-T H.264 et MPEG-2 et diverses erreurs de transmission (par exemple erreurs sur les bits, perte de paquets). Le modèle décrit dans la présente Recommandation peut être utilisé pour contrôler la qualité de réseaux déployés et s'assurer que ces réseaux sont prêts à être exploités. Les effets visuels des dégradations peuvent comprendre des dégradations aussi bien spatiales que temporelles. Le modèle décrit dans la présente Recommandation peut également être utilisé pour des tests en laboratoire de systèmes vidéo. Lorsque ce modèle est utilisé pour comparer différents systèmes vidéo, il est recommandé d'utiliser une méthode quantitative (par exemple celle de la Recommandation UIT-T J.149) pour déterminer la précision du modèle dans ce contexte particulier.

La présente Recommandation est censée être applicable aux services de radiodiffusion fournis à un débit compris entre 1 Mbit/s et 30 Mbit/s. Les résolutions et fréquences d'image ci-après ont été utilisées dans les tests de validation:

- 1080/59,94/I
- 1080/25/P
- 1080/50/I
- 1080/29,97/P.

Les tests de validation ont été effectués, pour chaque résolution, dans les conditions ci-après:

Facteurs de test
Résolution vidéo: 1920 × 1080 entrelacé et progressif
Fréquences d'image: 29,97 et 25 images par seconde
Débits binaires vidéo: 1 à 30 Mbit/s
Gel d'image temporel (pause avec saut) de 2 secondes au maximum
Erreurs de transmission avec perte de paquets
Conversion du circuit SRC de 1080 à 720/P, compression, transmission, décompression, puis nouvelle conversion à 1080
Technologies de codage
H.264/AVC (MPEG-4 Partie 10)
MPEG-2

Il est à noter que la résolution 720/P a été utilisée dans le plan des tests de validation comme condition de test (HRC). Etant donné qu'à l'heure actuelle, la résolution 720/P fait généralement l'objet d'une conversion vers une résolution plus haute pour l'affichage, on a estimé que des circuits HRC 720/P seraient mieux adaptés pour ce format.

1.1 Applications

Les applications du modèle d'évaluation décrit dans la présente Recommandation sont notamment les suivantes (la liste n'est pas exhaustive):

- 1) Contrôle de la qualité vidéo au récepteur lorsqu'on dispose de canaux latéraux.
- 2) Contrôle de la qualité vidéo à des noeuds de mesure situés entre le point d'émission et le point de réception.

Le modèle décrit dans la présente Recommandation offre une performance analogue statistiquement à celle du PSNR, mais il peut être utilisé pour l'évaluation de la qualité vidéo lorsqu'on dispose du signal de référence réduit au point de mesure.

1.2 Limitations

Le modèle d'évaluation de la qualité vidéo décrit dans la présente Recommandation ne peut pas être utilisé pour remplacer des tests subjectifs. Les valeurs de corrélation entre deux tests subjectifs conçus et exécutés avec soin (par exemple dans deux laboratoires différents) se situent normalement dans la fourchette 0,95-0,98. La présente Recommandation ne peut pas être utilisée pour comparer des systèmes vidéo (par exemple comparaison de deux codecs, comparaison de deux mises en oeuvre différentes du même algorithme de compression). La performance du modèle d'évaluation de la qualité vidéo décrit dans la présente Recommandation n'est pas meilleure statistiquement que celle du PSNR.

Lorsqu'un gel d'image était présent dans les conditions de test, sa durée était généralement inférieure à 2 secondes. Le modèle décrit dans la présente Recommandation n'a pas été validé pour mesurer la qualité vidéo dans une situation de remise en tampon (séquence vidéo avec un délai augmentant constamment ou gel d'image sans saut). Le modèle n'a pas été testé pour des fréquences d'image autres que celles utilisées dans les systèmes de télévision (à savoir 29,97 images par seconde et 25 images par seconde, en mode entrelacé ou progressif).

Il convient de noter que, dans le cas de nouvelles technologies de codage et de transmission introduisant des artéfacts qui n'ont pas été pris en compte dans la présente évaluation, le modèle de mesure objective pourrait donner des résultats erronés. Dans ce cas, une évaluation subjective est indispensable.

Il est à noter que le modèle décrit dans la présente Recommandation n'a pas été évalué pour un contenu de type «tête parlante», typique des scénarios de visioconférence.

2 Références

Les Recommandations UIT-T et autres références suivantes contiennent des dispositions qui, par suite de la référence qui y est faite, constituent des dispositions de la présente Recommandation. Au moment de la publication, les éditions indiquées étaient en vigueur. Les Recommandations et autres références étant sujettes à révision, les utilisateurs de la présente Recommandation sont invités à rechercher la possibilité d'appliquer les éditions les plus récentes des Recommandations et autres références énumérées ci-dessous. Une liste des Recommandations UIT-T en vigueur est publiée périodiquement.

La référence à un document figurant dans la présente Recommandation ne donne pas à ce document en tant que tel le statut de Recommandation.

Recommandation UIT-T J.244 (2008), *Méthodes d'étalonnage du désalignement constant des domaines spatial et temporel avec un gain et un décalage constants*.

3 Définitions

3.1 Termes définis ailleurs

La présente Recommandation utilise les termes suivants définis ailleurs:

3.1.1 Evaluation subjective (image): détermination de la qualité ou de la dégradation d'images de type «programme de télévision» présentées à un groupe d'évaluateurs pendant des séances de visionnement.

3.1.2 Mesure objective de la qualité perçue (image): mesure de la qualité de fonctionnement d'une chaîne de programme par l'emploi d'images de type «programme de télévision» et de méthodes de mesure objective (au moyen d'instruments) pour obtenir une indication proche de la note qui serait obtenue au moyen d'une évaluation subjective.

3.1.3 Proposant: organisation ou entreprise qui propose de soumettre un modèle de mesure de la qualité vidéo à des tests de validation, en vue de son inclusion éventuelle dans une Recommandation de l'UIT.

3.2 Termes définis dans la présente Recommandation

La présente Recommandation définit les termes suivants:

3.2.1 Fréquence d'image: nombre d'images uniques (nombre total d'images – nombre d'images répétées) par seconde.

3.2.2 Erreurs de transmission simulées: erreurs imposées au flux binaire vidéo numérique dans un environnement très contrôlé. On citera à titre d'exemple les taux de perte de paquets simulés et les erreurs sur les bits simulées. Les paramètres utilisés pour contrôler les erreurs de transmission simulées sont bien définis.

3.2.3 Erreurs de transmission: toute erreur imposée à la transmission vidéo. Comme exemples de type d'erreur, on citera les erreurs de transmission simulées et les conditions de réseau en exploitation.

4 Abréviations et acronymes

La présente Recommandation utilise les abréviations et acronymes suivants:

ACR	évaluation par catégories absolues (<i>absolute category rating</i>) (voir la Recommandation UIT-R BT.500)
ACR-HR	évaluation par catégories absolues avec référence cachée (<i>absolute category rating with hidden reference</i>) (voir la Recommandation UIT-T P.910)
AVI	entrelacement audio vidéo (<i>audio video interleave</i>)
DMOS	note d'opinion moyenne différentielle (<i>difference mean opinion score</i>)
FR	référence complète (<i>full reference</i>)
FRTV	télévision avec référence complète (<i>full reference television</i>)
HRC	circuit de référence fictif (<i>hypothetical reference circuit</i>)

ILG	groupe de laboratoires indépendants (<i>independent laboratory group</i>) du VQEG
MOS	note d'opinion moyenne (<i>mean opinion score</i>)
MOSp	note d'opinion moyenne prédite (<i>mean opinion score, predicted</i>)
NR	pas de référence ou référence nulle (<i>no (or zero) reference</i>)
PSNR	rapport signal de crête sur bruit (<i>peak signal-to-noise ratio</i>)
PVS	séquence vidéo traitée (<i>processed video sequence</i>)
RMSE	erreur quadratique moyenne (<i>root mean square error</i>)
RR	référence réduite (<i>reduced reference</i>)
SFR	fréquence d'image de la source (<i>source frame rate</i>)
SRC	canal ou circuit de référence de la source (<i>source reference channel or circuit</i>)
VQEG	groupe d'experts en qualité vidéo (<i>video quality experts group</i>)
YUV	espace chromatique et format de fichier

5 Conventions

Aucune.

6 Description de la méthode de mesure avec référence réduite

6.1 Introduction

Bien que le rapport PSNR soit largement utilisé pour la mesure objective de la qualité vidéo, on signale qu'il ne représente pas bien la qualité vidéo perçue. En analysant comment l'être humain perçoit la qualité vidéo, on constate que le système visuel humain est sensible aux dégradations sur les contours. En d'autres termes, lorsque les pixels des contours d'une séquence vidéo sont flous, les évaluateurs ont tendance à donner des notes basses à la séquence vidéo même si le PSNR est élevé. A partir de ce constat, on a mis au point des modèles avec référence réduite qui mesurent principalement les dégradations sur les contours.

La Fig. 1 illustre le fonctionnement d'un modèle avec référence réduite. Les caractéristiques qui seront utilisées pour mesurer la qualité vidéo en un point de contrôle sont extraites de la séquence vidéo source et transmises. Le Tableau 1 donne les largeurs de bande des canaux latéraux pour les caractéristiques, qui ont été utilisées dans le test sur la TVHD mené par le Groupe VQEG.

FIGURE 1

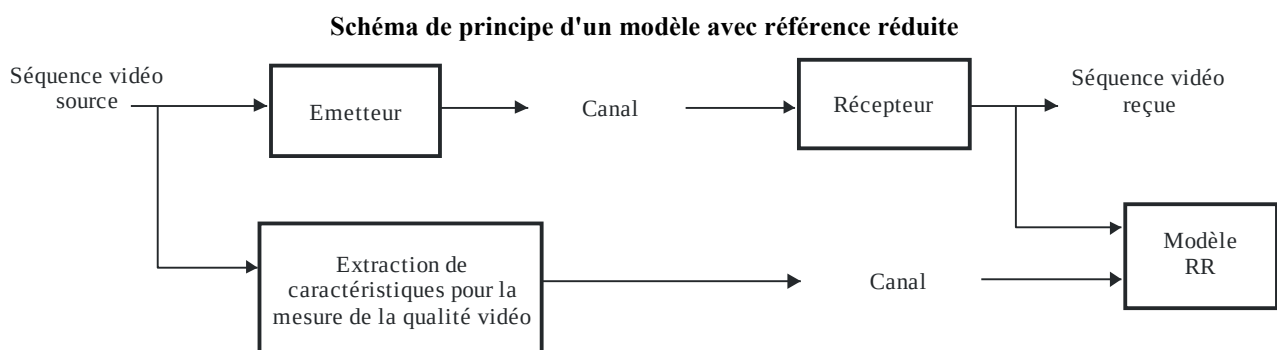


TABLEAU 1

Largeurs de bande des canaux latéraux

Format vidéo	Largeurs de bandes utilisées dans les tests
1080/60 Hz (29,97 images/s) 1080/30Pp (29,97 images/s)	56 kbit/s, 128 kbit/s, 256 kbit/s
1080/25Pp (25 images/s) 1080/50I Hz (25 images/s)	56 kbit/s, 128 kbit/s, 256 kbit/s

6.2 Modèle avec référence réduite EPSNR

6.2.1 EPSNR

Les modèles avec référence réduite mesurent principalement les dégradations sur les contours. Dans les modèles, on commence par appliquer un algorithme de détection des contours à la séquence vidéo source afin de localiser les pixels des contours. On mesure ensuite la dégradation de ces pixels en calculant l'erreur quadratique moyenne. Puis, à partir de cette erreur quadratique moyenne, on calcule le PSNR pour les contours (EPSNR, *edge PSNR*).

On peut utiliser n'importe quel algorithme de détection des contours, même si les résultats peuvent être légèrement différents. On peut par exemple utiliser un opérateur de gradient pour localiser les pixels des contours. Plusieurs opérateurs de gradient ont été proposés. Dans de nombreux algorithmes de détection des contours, on commence par calculer l'image des gradients horizontaux $g_{horizontal}(m,n)$ et l'image des gradients verticaux $g_{vertical}(m,n)$ au moyen d'opérateurs de gradient. On peut alors calculer l'image des gradients d'amplitude $g(m,n)$ comme suit:

$$g(m,n) = |g_{horizontal}(m,n)| + |g_{vertical}(m,n)|$$

Enfin, on applique une opération de seuillage à l'image des gradients d'amplitude pour trouver les pixels des contours. En d'autres termes, les pixels dont les gradients d'amplitude sont supérieurs à une valeur seuil sont considérés comme étant des pixels de contour.

Les Fig. 2 à 6 illustrent la procédure. La Fig. 2 montre une image source. La Fig. 3 montre une image de gradients horizontaux $g_{horizontal}(m,n)$, obtenue par l'application d'un opérateur de gradient horizontal à l'image source de la Fig. 2. La Fig. 4 montre une image de gradients verticaux $g_{vertical}(m,n)$, obtenue par l'application d'un opérateur de gradient vertical à l'image source de la Fig. 2. La Fig. 5 montre l'image des gradients d'amplitude (image des contours) et la Fig. 6 montre l'image binaire des contours (image masque) obtenue par l'application d'un seuillage à l'image des gradients d'amplitude de la Fig. 5.

FIGURE 2

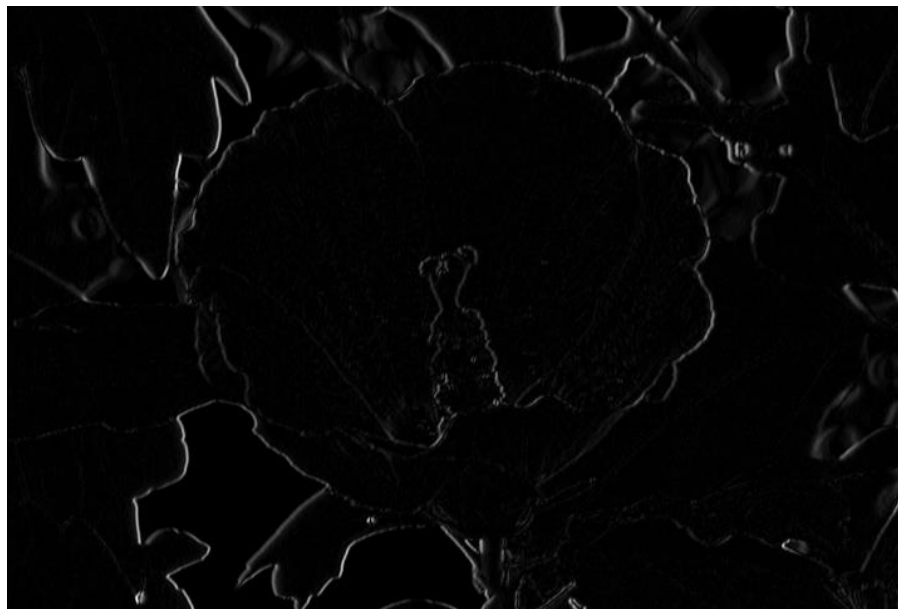
Image source (image d'origine)



BT.1908-02

FIGURE 3

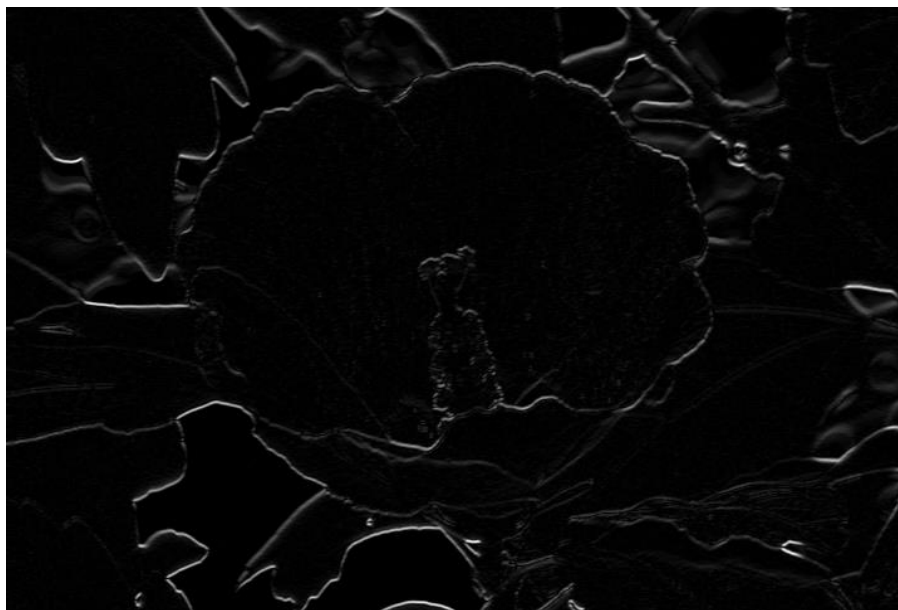
Image de gradients horizontaux, obtenue par l'application d'un opérateur de gradient horizontal à l'image source de la Fig. 2



BT.1908-03

FIGURE 4

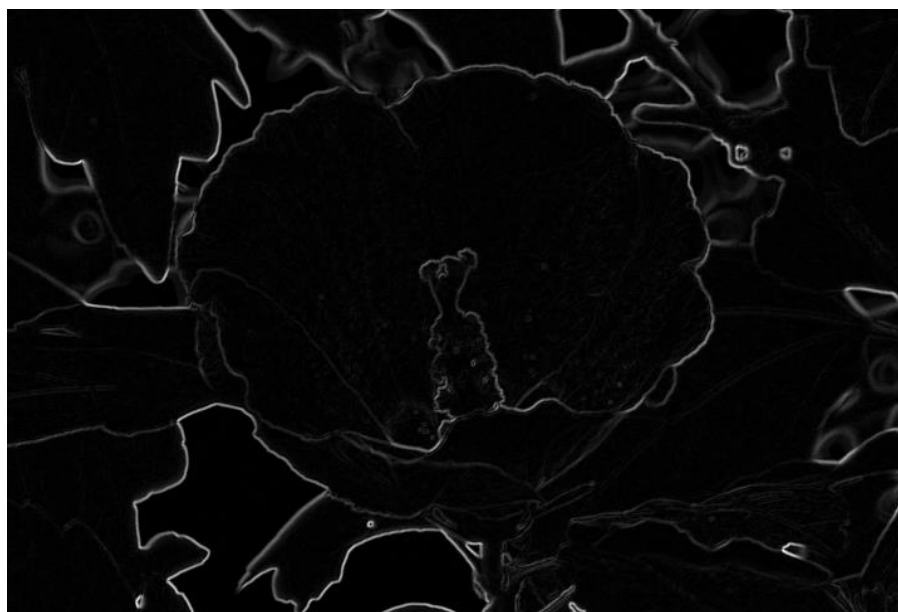
Image de gradients verticaux, obtenue par l'application d'un opérateur de gradient vertical à l'image source de la Fig. 2



BT.1908-04

FIGURE 5

Image des gradients d'amplitude



BT.1908-05

FIGURE 6

Image binaire des contours (image masque) obtenue par l'application d'un seuillage à l'image des gradients d'amplitude de la Fig. 5

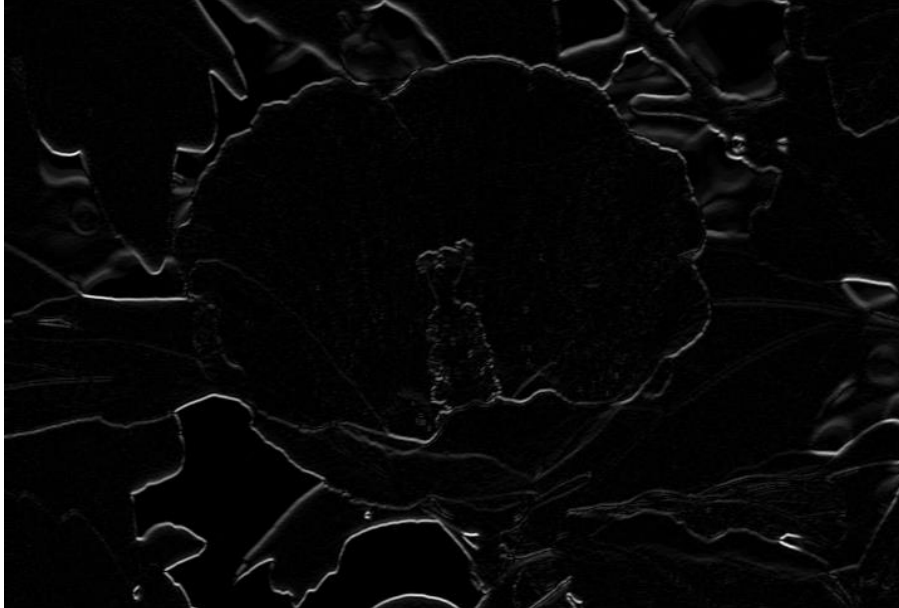


BT.1908-06

On peut également utiliser une procédure modifiée pour trouver les pixels de contour. Par exemple, on peut tout d'abord appliquer un opérateur de gradient vertical à l'image source, ce qui donne une image de gradients verticaux. On applique ensuite un opérateur de gradient horizontal à l'image de gradients verticaux, ce qui donne une image de gradients successifs modifiée (image de gradients horizontaux et verticaux). On peut alors appliquer un seuillage à l'image de gradients successifs modifiée pour trouver les pixels des contours. En d'autres termes, les pixels de l'image de gradients successifs modifiée qui dépassent une valeur seuil sont considérés comme des pixels de contour. Les Fig. 7 à 9 illustrent la procédure modifiée. La Fig. 7 montre une image de gradients verticaux $g_{vertical}(m,n)$, obtenue par l'application d'un opérateur de gradient vertical à l'image source de la Fig. 2. La Fig. 8 montre une image de gradients successifs modifiée (image de gradients horizontaux et verticaux), obtenue par l'application d'un opérateur de gradient horizontal à l'image de gradients verticaux de la Fig. 7. La Fig. 9 montre l'image binaire des contours (image masque) obtenue par l'application d'un seuillage à l'image de gradients successifs modifiée de la Fig. 8.

FIGURE 7

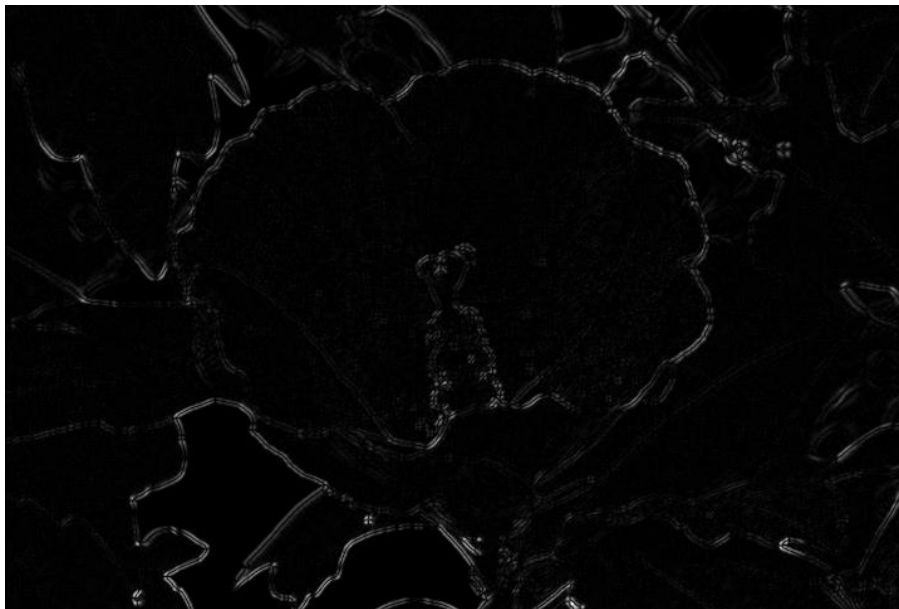
Image de gradients verticaux, obtenue par l'application d'un opérateur de gradient vertical à l'image source de la Fig. 2



BT.1908-07

FIGURE 8

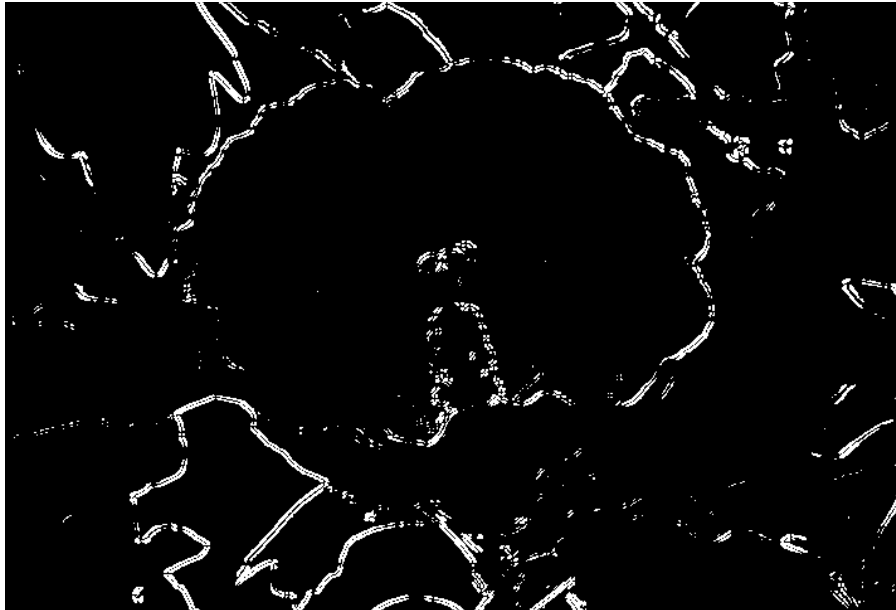
Image de gradients successifs modifiée (image de gradients horizontaux et verticaux), obtenue par l'application d'un opérateur de gradient horizontal à l'image de gradients verticaux de la Fig. 7



BT.1908-08

FIGURE 9

Image binaire des contours (image masque) obtenue par l'application d'un seuillage à l'image de gradients successifs modifiée de la Fig. 8



BT.1908-09

Il est à noter que les deux méthodes peuvent être considérées comme étant des algorithmes de détection des contours. On peut choisir tout algorithme de détection des contours selon la nature des séquences vidéo et les algorithmes de compression. Toutefois, certaines méthodes peuvent donner de meilleurs résultats que d'autres.

Ainsi, dans le modèle, on applique tout d'abord un opérateur de détection des contours afin d'obtenir une image des contours (Fig. 5 et 8). Puis, on obtient une image de masque (image binaire des contours) en appliquant un seuillage à l'image des contours (Fig. 6 et 9). En d'autres termes, les pixels de l'image des contours dont la valeur est inférieure au seuil t_e sont mis à zéro et les pixels dont la valeur est égale ou supérieure au seuil sont mis à une valeur non nulle. Les Fig. 6 et 9 montrent des images de masque. Etant donné qu'une séquence vidéo peut être considérée comme une séquence d'images ou de trames, la procédure précitée peut être appliquée à chaque image ou trame des séquences vidéo. Puisque le modèle peut être utilisé pour les séquences vidéo à trames ou à images, le terme «image» sera utilisé pour désigner indifféremment une image ou une trame.

6.2.2 Extraction de caractéristiques des séquences vidéo source

Etant donné que le modèle est un modèle avec référence réduite, un ensemble de caractéristiques doit être extrait de chaque image d'une séquence vidéo source. Dans le modèle avec référence réduite EPSNR, un certain nombre de pixels de contour est extrait de chaque image. Les positions et les valeurs de ces pixels sont ensuite codées et transmises. Toutefois, pour certaines séquences vidéo, le nombre de pixels de contour peut être très faible lorsqu'on utilise une valeur seuil fixe. Dans le scénario le plus défavorable, ce nombre peut être nul (images vides ou images à cadence très faible). Afin de résoudre ce problème, si le nombre de pixels de contour d'une image est inférieur à une valeur donnée, l'utilisateur peut abaisser le seuil jusqu'à ce que le nombre de pixels de contour soit supérieur à une certaine valeur. On peut aussi choisir comme pixels de contour les pixels qui correspondent aux plus grandes valeurs de l'image de gradients horizontaux et verticaux. Lorsqu'il n'y a pas de pixels de contour (image vide par exemple) dans une image, on peut choisir aléatoirement le nombre de pixels requis ou sauter l'image. Par exemple, si dix pixels de contour doivent être choisis dans chaque image, on peut trier les pixels de l'image de gradients horizontaux

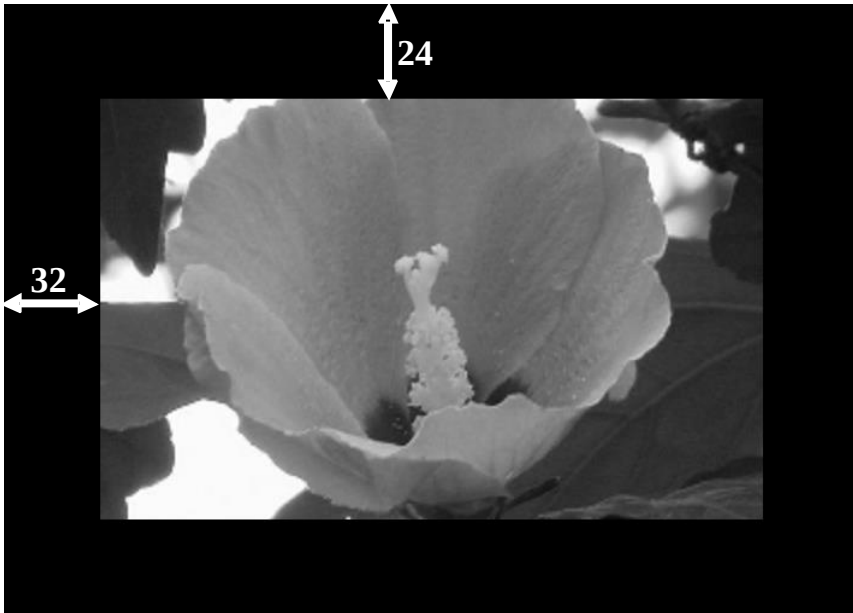
et verticaux selon leurs valeurs et retenir les dix plus grandes valeurs. Toutefois, cette procédure peut donner plusieurs pixels de contour à des positions identiques. Pour résoudre ce problème, on peut d'abord choisir plusieurs fois le nombre souhaité de pixels de l'image de gradients horizontaux et verticaux et choisir ensuite aléatoirement le nombre souhaité de pixels de contour parmi les pixels choisis de l'image de gradients horizontaux et verticaux. Dans les modèles testés lors du test sur la TVHD mené par le Groupe VQEG, le nombre souhaité de pixels de contour est choisi aléatoirement parmi un grand ensemble de pixels de contour, obtenu par l'application d'un seuillage à l'image de gradients.

Dans les modèles avec référence réduite EPSNR, les positions et les valeurs des pixels de contour sont codées après l'application d'un filtre passe-bas gaussien aux emplacements de pixels choisis. Bien que l'on ait utilisé le filtre passe-bas gaussien (7×3) dans le test sur la TVHD mené par le Groupe VQEG, des filtres passe-bas différents peuvent être utilisés suivant les formats vidéo. Il est à noter que pendant le processus de codage, on peut appliquer un recadrage. Afin d'éviter de choisir des pixels de contour dans les zones recadrées, le modèle choisit les pixels de contour dans la zone centrale (Fig. 10). Le Tableau 2 indique les dimensions après recadrage. Il indique aussi le nombre de bits nécessaires pour coder la position et la valeur d'un pixel de contour.

TABLEAU 2
Bits nécessaires par pixel de contour

Format vidéo	Dimensions	Dimensions après recadrage	Bits de position	Bits de valeur de pixel	Nbre total de bits par pixel
HD progressif	1 920 × 1 080	1 856 × 1032	21	8	29
HD entrelacé	1 920 × 540	1 856 × 516	20	8	28

FIGURE 10
Exemple de recadrage et zone centrale

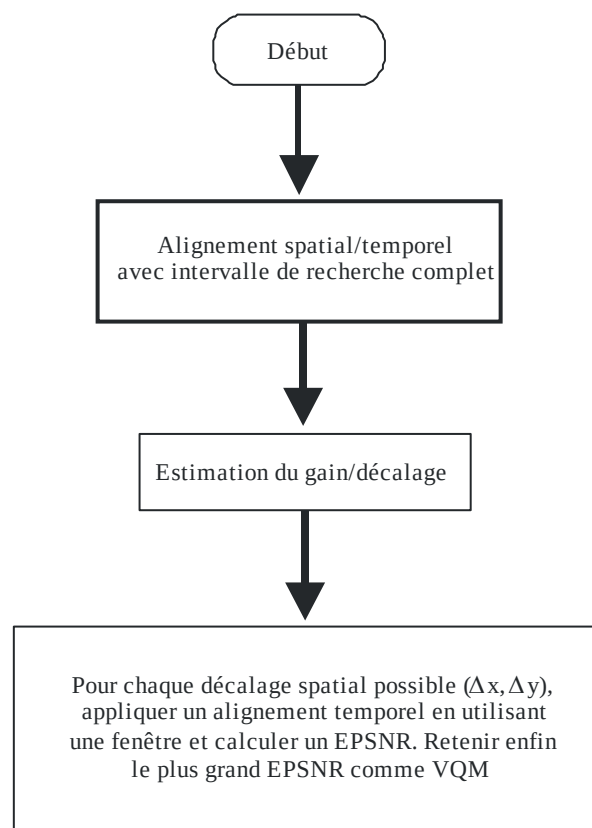


Le modèle choisit les pixels de contour dans chaque image conformément à la largeur de bande autorisée (Tableau 1). Le Tableau 3 indique le nombre de pixels de contour par image qui peut être transmis pour les largeurs de bande utilisées dans les tests.

TABLEAU 3
Nombre de pixels de contour par image/trame

Format vidéo	56 kbit/s	128 kbit/s	256 kbit/s
HD progressif	46	105	211
HD entrelacé	24	54	109

FIGURE 11
Organigramme du modèle



BT.1908-11

6.2.3 Alignement spatial/temporel et réglage du gain/décalage

Avant de calculer la différence entre les pixels de contour de la séquence vidéo source et ceux de la séquence vidéo traitée, qui est la séquence vidéo reçue au récepteur, le modèle procède d'abord à un alignement spatial/temporel et à un réglage du gain/décalage. On a utilisé la méthode d'étalonnage (Annexe B) de la Recommandation UIT-T J.244. Pour transmettre les caractéristiques de gain et de décalage de cette Recommandation (Annexe B), on a utilisé 30% de la largeur de bande disponible dans le test sur la TVHD mené par le Groupe VQEG. Pour une séquence vidéo entrelacée, la méthode d'étalonnage est appliquée trois fois (trames paires, trames impaires et images combinées) tandis que pour les séquences vidéo progressives, elle est appliquée aux images. Si la différence entre l'erreur de trame paire (PSNR) et l'erreur de trame impaire est supérieure à un seuil, les

résultats de l'alignement (décalage en/x, décalage en/y) avec la plus petite valeur de PSNR ont été utilisés. Dans le cas contraire, on a utilisé les résultats de l'alignement avec les images combinées. Dans le test sur la TVHD mené par le Groupe VQEG, le seuil a été fixé à 2 dB.

Au point de contrôle, la séquence vidéo traitée doit être alignée à l'aide des pixels de contour extraits de la séquence vidéo source. Toutefois, si la largeur de bande des canaux latéraux est petite, seuls quelques pixels de contour de la séquence vidéo source sont disponibles (Fig. 12). En conséquence, l'alignement temporel peut être imprécis s'il est effectué en utilisant une seule image (Fig. 13). Pour résoudre ce problème, le modèle fait appel à une fenêtre pour l'alignement temporel. Au lieu d'utiliser une seule image de la séquence vidéo traitée, le modèle construit une fenêtre qui est composée de plusieurs images adjacentes pour trouver le décalage temporel optimal. La Fig. 14 illustre la procédure. L'erreur quadratique moyenne dans la fenêtre est calculée comme suit:

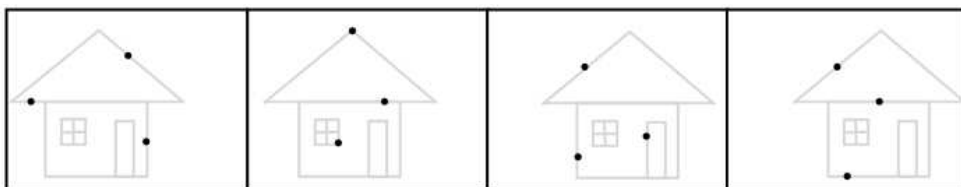
$$MSE_{window} = \frac{1}{N_{win}} \sum (E_{SRC}(i) - E_{PVS}(i))^2$$

où MSE_{window} est l'erreur quadratique moyenne dans la fenêtre, $E_{SRC}(i)$ est un pixel de contour à l'intérieur de la fenêtre qui a un pixel correspondant dans la séquence vidéo traitée, $E_{PVS}(i)$ est un pixel de la séquence vidéo traitée correspondant au pixel de contour et N_{win} est le nombre total de pixels de contour utilisé pour calculer MSE_{window} . Cette erreur quadratique moyenne dans la fenêtre est utilisée comme différence entre une image de la séquence vidéo traitée et l'image correspondante de la séquence vidéo source.

La taille de la fenêtre peut être déterminée en prenant en considération la nature de la séquence vidéo traitée. Pour une application type, une fenêtre correspondant à deux secondes est recommandée. On peut aussi appliquer différentes tailles de fenêtre et utiliser la meilleure, c'est-à-dire celle qui donne la plus faible erreur quadratique moyenne. En outre, on peut utiliser différents centres de fenêtre pour tenir compte du saut d'images dû à des erreurs de transmission (Fig. 18).

FIGURE 12

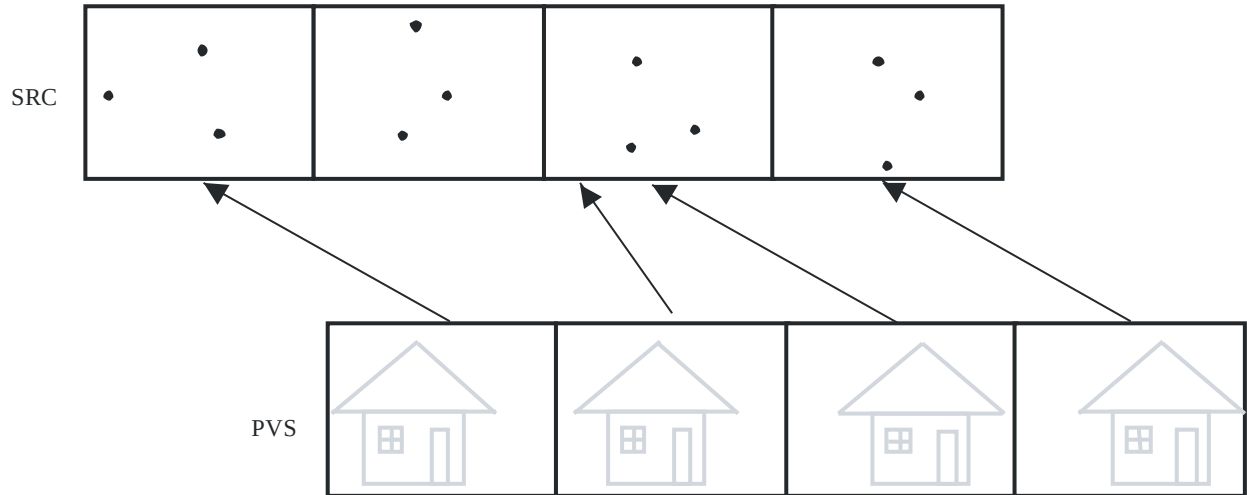
Choix des pixels de contour dans une séquence vidéo source



BT.1908-12

FIGURE 13

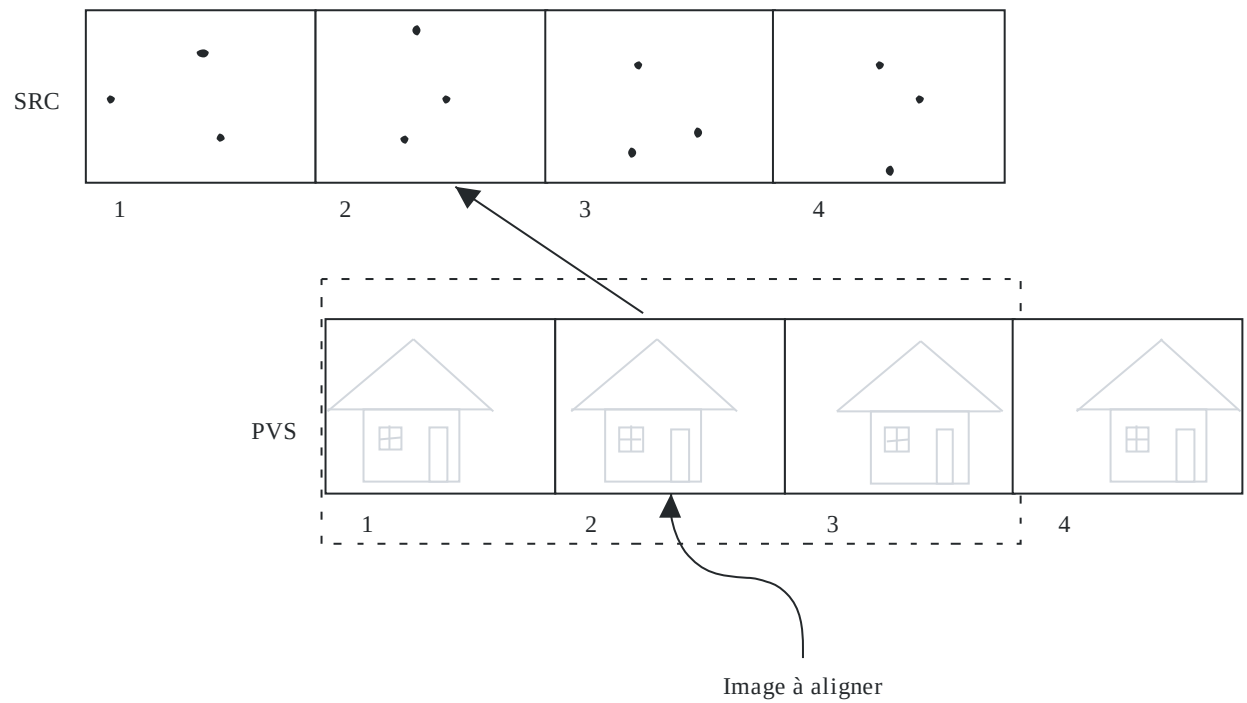
Alignement de la séquence vidéo traitée au moyen des pixels de contour de la séquence vidéo source



BT.1908-13

FIGURE 14

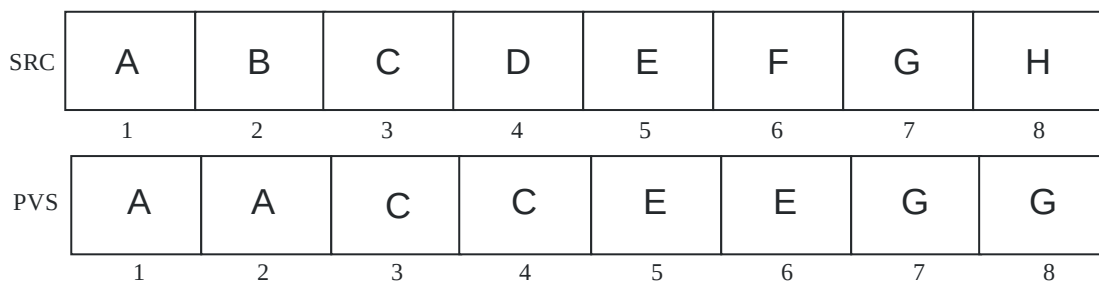
Alignement de la séquence vidéo traitée au moyen des pixels de contour à l'intérieur d'une fenêtre



BT.1908-14

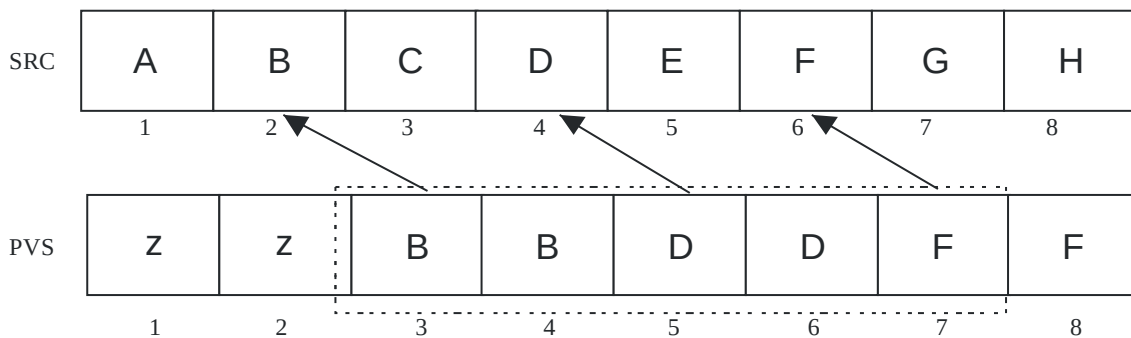
Lorsque la séquence vidéo source est codée avec des taux de compression élevés, le codeur peut réduire le nombre d'images par seconde et la séquence vidéo traitée comporte des images répétées (Fig. 15). Sur la Fig. 15, la séquence vidéo traitée ne comporte pas d'images correspondant à certaines images de la séquence vidéo source (2ème, 4ème, 6ème et 8ème images). Dans ce cas, le modèle n'utilise pas les images répétées dans le calcul de l'erreur quadratique moyenne. En d'autres termes, le modèle effectue un alignement temporel en utilisant la première image (image valable) de chaque bloc répété. Ainsi, sur la Fig.16, seules trois images (3ème, 5ème et 7ème images) dans la fenêtre sont utilisées pour l'alignement temporel.

FIGURE 15

Exemple d'images répétées

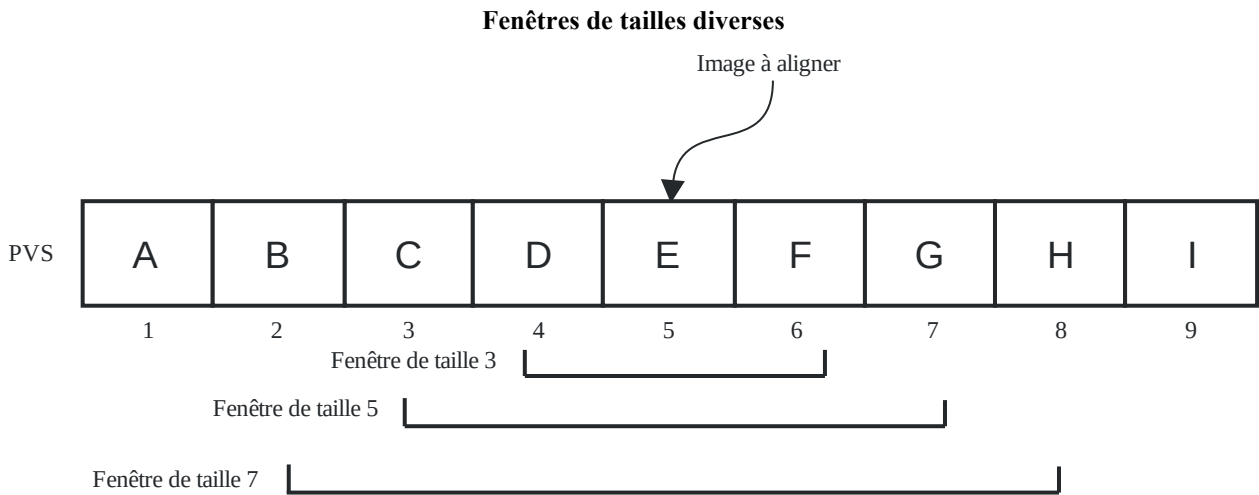
BT.1908-15

FIGURE 16

Traitement des images répétées

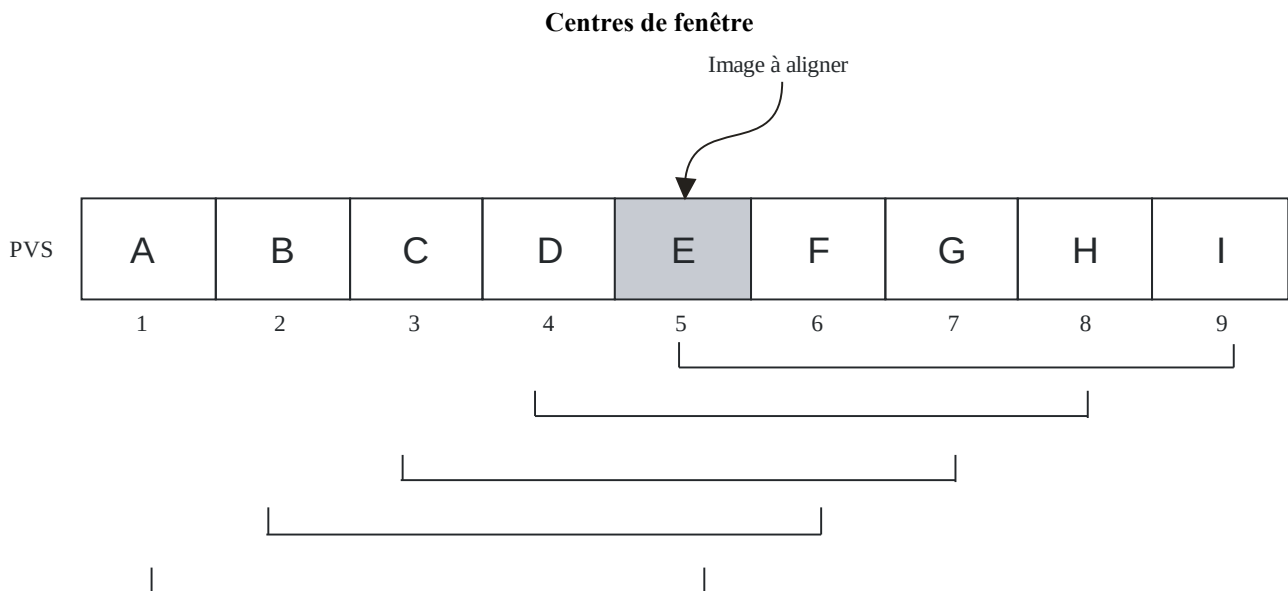
BT.1908-16

FIGURE 17



BT.1908-17

FIGURE 18



BT.1908-18

6.2.4 Calcul de l'EPSNR et post-traitement

Après l'alignement temporel, on calcule la moyenne des différences entre les pixels de contour de la séquence vidéo source et les pixels correspondants de la séquence vidéo traitée, moyenne qui peut être interprétée comme correspondant à l'erreur quadratique moyenne des contours de la séquence vidéo traitée (MSE_{edge}). On calcule enfin l'EPSNR (PSNR pour les contours) comme suit:

$$EPSNR = 10 \log_{10} \left(\frac{P^2}{MSE_{edge}} \right)$$

où p est la valeur de crête de l'image.

Etant donné que diverses dégradations peuvent réduire la qualité vidéo, la valeur d'EPSNR est ajustée compte tenu des effets quantifiés dans les sous-paragraphes qui suivent.

1) Paramètre I d'effet de blocs

Pour tenir compte de l'effet de blocs, on calcule les différences moyennes entre les colonnes. En prenant pour hypothèse le modulo 8, la note d'effet de blocs pour la i ème image est calculée comme suit:

$$Blk[i] = \frac{\text{plus grande différence entre les colonnes}}{\text{deuxième plus grande différence entre les colonnes}}$$

La note finale d'effet de blocs (*Blocking*) est calculée en faisant la moyenne des notes d'effet de blocs pour chaque image:

$$Blocking = \frac{1}{\text{number of frames}} \sum_i Blk[i]$$

Enfin, on utilise les formules suivantes:

IF(BLOCKING > 12 and $25 \leq EPSNR < 30$) adjust_EPSNR_blk1=3

IF(BLOCKING > 5 and $30 \leq EPSNR < 35$) adjust_EPSNR_blk1=5

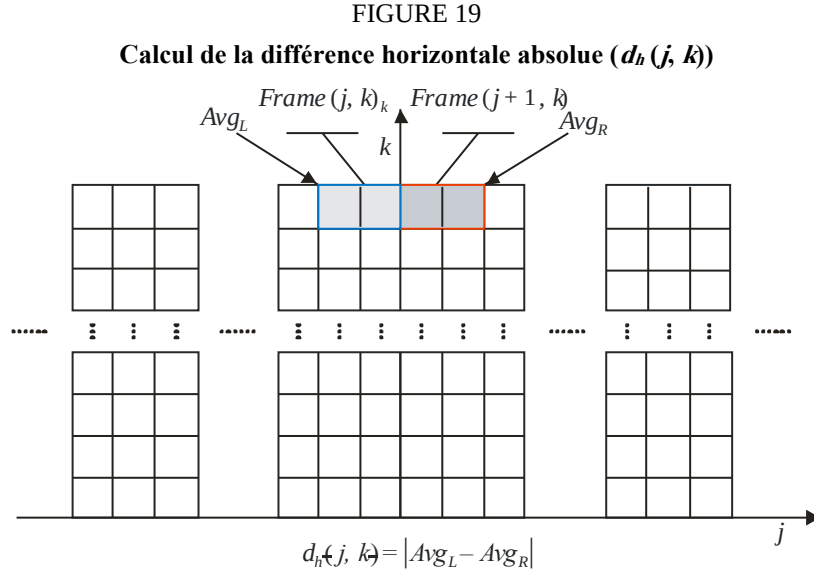
2) Paramètre II d'effet de blocs

Dans l'hypothèse où des dégradations dues à l'effet de blocs sont susceptibles de se produire dans chaque 8ème colonne (par exemple en MPEG2), on utilise également un deuxième paramètre d'effet de blocs. Pour calculer ce paramètre, on commence par calculer la différence horizontale absolue comme suit (Fig. 19):

$$d_h(j, k) = |Avg_L - Avg_R|$$

où:

$$Avg_L = \frac{1}{2} \sum_{p=-1}^0 Frame(j+p, k), \quad Avg_R = \frac{1}{2} \sum_{p=1}^2 Frame(j+p, k)$$



BT.1908-19

La somme des effets de blocs horizontaux (SB_h) à la position j est alors définie comme suit:

$$SB_h[j] = \left(\sum_{1 \leq k \leq height} (|Frame(j, k) - Frame(j+1, k)| \times u(d_h(j, k) - \Phi(Avg_L))) \right)^2$$

où $u(\cdot)$ représente la fonction échelon unité et:

$$\Phi(s) = \begin{cases} 17(1 - \sqrt{s/127}) + 3 & \text{si } s \leq 127 \\ 3(s - 127)/128 + 3 & \text{sinon} \end{cases}$$

La procédure est répétée pour la totalité de l'image, puis on calcule la caractéristique d'effet de blocs horizontal pour l'image (FB_h) comme suit:

$$FB_h = \left(\sum_{\substack{1 \leq j \leq width \\ j \equiv 0 \pmod{8}}} SB_h(j) \right)^{1/2}$$

Pour chaque image, la différence entre colonnes (NFB_h), exception faite de chaque 8ème colonne, est calculée comme suit:

$$NFB_h = \frac{1}{7} \sum_{l=1}^7 \left(\sum_{\substack{1 \leq j \leq width \\ j \equiv l \pmod{8}}} \left(\sum_{1 \leq k \leq height} (|Frame(j, k) - Frame(j+1, k)| \times u(d_h(j, k) - \Phi(Avg_L))) \right)^2 \right)^{1/2}$$

La caractéristique d'effet de blocs horizontal finale, BLK_H , est ensuite calculée comme suit:

$$BLK_H = \ln(FB_h / NFB_h)$$

La caractéristique d'effet de blocs vertical BLK_V est calculée de manière analogue. Pour les séquences vidéo entrelacées, la caractéristique d'effet de blocs vertical est calculée dans la séquence de trames. La note d'effet de blocs pour la i ème image est calculée comme suit:

$$FrameBLK(i) = 0,5 \times BLK_H + 0,5 \times BLK_V$$

On calcule la note finale d'effet de blocs (*BLOCKING2*) en faisant la moyenne sur les 10% supérieurs des notes d'effet de blocs pour chaque image.

Enfin, on utilise les formules suivantes:

IF(<i>BLOCKING2</i> > 1.5 and $25 \leq \text{EPSNR} < 30$)	adjust_EPSNR_blk2=2
IF(<i>BLOCKING2</i> > 1.3 and $30 \leq \text{EPSNR} < 35$)	adjust_EPSNR_blk2=2
IF(<i>BLOCKING2</i> > 1.5 and $35 \leq \text{EPSNR} < 40$)	adjust_EPSNR_blk2=2
IF(<i>BLOCKING2</i> > 1 and $40 \leq \text{EPSNR} < 45$)	adjust_EPSNR_blk2=2
IF(<i>BLOCKING2</i> > 0.5 and $45 \leq \text{EPSNR} < 55$)	adjust_EPSNR_blk2=2

Comme on peut le voir dans les formules ci-dessus, cet ajustement a peu d'effets sur la valeur finale d'EPSNR. En l'absence d'artéfacts de type effet de blocs dans chaque 8ème colonne, on peut sauter cet ajustement ou commencer par trouver les emplacements présentant un effet de blocs. On peut aussi utiliser une fonction différente pour $\Phi(s)$.

3) Durée maximale de gel d'une image et durée totale de gel des images

Les erreurs de transmission peuvent avoir pour conséquence des images gelées pendant une longue durée. Pour tenir compte de ce phénomène, on utilise les formules suivantes:

IF(<i>MAX_FREEZE</i> ≥ 8 and $25 \leq \text{EPSNR} < 30$)	adjust_EPSNR_max_freeze=3
IF(<i>MAX_FREEZE</i> ≥ 6 and $30 \leq \text{EPSNR} < 35$)	adjust_EPSNR_max_freeze=3
IF(<i>MAX_FREEZE</i> ≥ 3 and $35 \leq \text{EPSNR} < 40$)	adjust_EPSNR_max_freeze=3
IF(<i>MAX_FREEZE</i> ≥ 1.5 and $40 \leq \text{EPSNR} < 45$)	adjust_EPSNR_max_freeze=2
IF(<i>MAX_FREEZE</i> ≥ 1 and $45 \leq \text{EPSNR} < 95$)	adjust_EPSNR_max_freeze=2

où *MAX_FREEZE* est la plus longue durée de gel d'une image. Il est à noter que si la séquence vidéo n'est pas de 10 s, il conviendra d'utiliser des seuils différents.

Il est également tenu compte de la durée totale de gel des images comme suit:

IF(<i>TOTAL_FREEZE</i> ≥ 80 and $25 \leq \text{EPSNR} < 30$)	adjust_EPSNR_total_freeze=3
IF(<i>TOTAL_FREEZE</i> ≥ 40 and $30 \leq \text{EPSNR} < 35$)	adjust_EPSNR_total_freeze=4
IF(<i>TOTAL_FREEZE</i> ≥ 10 and $35 \leq \text{EPSNR} < 40$)	adjust_EPSNR_total_freeze=3.5
IF(<i>TOTAL_FREEZE</i> ≥ 2 and $\text{EPSNR} \geq 40$)	adjust_EPSNR_total_freeze=1.5

où *TOTAL_FREEZE* est la durée totale de gel des images. Il est à noter que si la séquence vidéo n'est pas de 10 s, il conviendra d'utiliser des seuils différents.

4) Bloc d'erreurs de transmission

Des erreurs de transmission peuvent entraîner le gel de blocs locaux. En outre, dans les scènes statiques, certains blocs sont identiques aux blocs des images précédentes aux mêmes positions. Pour tenir compte du gel de blocs locaux dû à des erreurs de transmission, les blocs contenant les pixels de contour transmis sont classés comme étant soit des blocs identiques (les blocs sont identiques aux blocs des images précédentes) soit des blocs différents. On calcule ensuite deux EPSNR pour les blocs identiques et les blocs différents. Si la différence entre les deux EPSNR (*EPSNR_diff*) est grande, cela indique que des erreurs de transmission ont pu se produire. A partir de ce constat, on ajuste l'EPSNR comme suit:

IF($8 \leq \text{EPSNR_diff} \leq 30$ and $25 \leq \text{EPSNR} < 30$)	adjust_EPSNR_diff= 3
IF($9 \leq \text{EPSNR_diff} \leq 30$ and $30 \leq \text{EPSNR} < 35$)	adjust_EPSNR_diff= 4
IF($10 \leq \text{EPSNR_diff} \leq 30$ and $35 \leq \text{EPSNR} < 40$)	adjust_EPSNR_diff= 6
IF($9 \leq \text{EPSNR_diff} < 10$ and $35 \leq \text{EPSNR} < 40$)	adjust_EPSNR_diff= 2
IF($9 \leq \text{EPSNR_diff} \leq 30$ and $40 \leq \text{EPSNR} < 45$)	adjust_EPSNR_diff= 4

Toutefois, si le nombre total des blocs identiques est inférieur à 100, aucun ajustement n'est effectué.

5) Ajustement final de l'EPSNR

Enfin, la valeur d'EPSNR est ajustée comme suit:

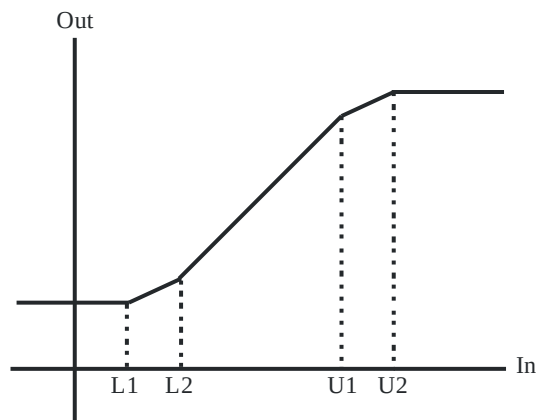
$$\text{EPSNR} \leq \text{EPSNR} - \text{MAX}(\text{adjust_EPSNR_blk1}, \text{adjust_EPSNR_blk2}, \text{adjust_EPSNR_max_freeze}, \text{adjust_EPSNR_total_freeze}, \text{adjust_EPSNR_diff})$$

6) Fonction linéaire par morceaux

Lorsque l'EPSNR est supérieur à une certaine valeur, la qualité perçue atteint un plafond. Dans ce cas, il est possible de fixer la limite supérieure de l'EPSNR. De plus, lorsqu'une relation linéaire entre l'EPSNR et la note d'opinion moyenne différentielle (DMOS, *difference mean opinion score*) est souhaitable, on peut appliquer une fonction linéaire par morceaux comme illustré sur la Fig. 20. Dans le modèle testé lors du test sur la TVHD mené par le Groupe VQEG, la limite supérieure a été fixée à 50 et la limite inférieure à 19.

FIGURE 20

Fonction linéaire par morceaux pour la relation linéaire entre l'EPSNR et la DMOS



Appendice 1 (informatif)

Conclusions du Groupe d'experts en qualité vidéo (VQEG)

Un groupe informel, le Groupe d'experts en qualité vidéo (VQEG), fait des études sur les mesures de la qualité vidéo perçue et fait rapport aux Commissions d'études 9 et 12 de l'UIT-T. Le test de phase I du Groupe VQEG sur la télévision haute définition récemment mené à bien par le Groupe VQEG a permis d'évaluer la performance des algorithmes proposés de mesure de la qualité vidéo perçue avec référence réduite.

Les statistiques qui suivent sont extraites du rapport final sur la TVHD du Groupe VQEG. Il est à noter que ce rapport contient aussi d'autres informations: corrélation de Pearson et RMSE calculées dans des expériences individuelles, intervalles de confiance, tests de significativité statistique dans des expériences individuelles, analyse de sous-ensembles des données qui incluent des dégradations spécifiques (par exemple codage UIT-T H.264 uniquement), diagrammes de dispersion, et coefficients d'ajustement.

Analyse principale

La performance du modèle avec référence réduite retenu est récapitulée dans le Tableau 4. Le rapport PSNR est calculé conformément à la Recommandation UIT-T J.340 et inclus dans cette analyse à des fins de comparaison. «Superset RMSE» désigne le paramètre principal (RMSE) calculé dans le superensemble agrégé (les six expériences reportées sur une même échelle). «Top performing group total» désigne le nombre d'expériences (0 à 6) pour lesquelles ce modèle a été le modèle le plus performant ou statistiquement équivalent au modèle le plus performant. «Better than PSNR total» désigne le nombre d'expériences (0 à 6) pour lesquelles le modèle a été statistiquement meilleur que le PSNR. «Better Than Superset PSNR» indique si chaque modèle est statistiquement meilleur que le PSNR dans le superensemble agrégé. «Superset Correlation» désigne la corrélation de Pearson calculée dans le superensemble agrégé.

TABLEAU 4

Paramètre	PSNR	Yonsei56k	Yonsei128k	Yonsei256k
Superset RMSE	0,71	0,73	0,73	0,73
Top performing group total	6	4	4	4
Equivalent to or better than PSNR total	6	4	4	4
Equivalent to superset PSNR	Oui	Oui	Oui	Oui
Superset correlation	0,78	0,77	0,77	0,77

Etant donné que la performance du modèle est identique sur le plan statistique pour les trois largeurs de bande, il est recommandé d'utiliser ce modèle avec au moins une largeur de bande des canaux latéraux de 56 kbit/s.

Analyse secondaire

Le Tableau 5 donne la RMSE pour le modèle avec référence réduite, pour des subdivisions du superensemble en fonction du type de codage (UIT-T H.264 ou MPEG-2), de la présence d'erreurs de transmission ou de la question de savoir si le circuit HRC contenait uniquement des artéfacts liés au codage. Etant donné que les expériences n'ont pas été conçues de sorte que ces variables couvrent uniformément toute la plage de qualité, seule la RMSE est présentée pour ces subdivisions.

TABLEAU 5

Type de circuit HRC	PSNR	Yonsei56k	Yonsei128k	Yonsei256k
Codage H.264	0,75	0,65	0,65	0,65
Erreur H.264	0,67	0,86	0,85	0,86
Codage MPEG-2	0,78	0,81	0,81	0,80
Erreur MPEG-2	0,66	0,68	0,68	0,68
Codage	0,75	0,69	0,69	0,69
Erreur	0,67	0,79	0,78	0,79

Bibliographie

Recommandation UIT-T P.910 (2008), *Méthodes subjectives d'évaluation de la qualité vidéographique pour les applications multimédias*.

Recommandation UIT-T P.911 (1998), *Méthodes d'évaluation subjective de la qualité audiovisuelle pour applications multimédias*.

Recommandation UIT-T J.143 (2000), *Prescriptions d'utilisateur relatives aux mesures objectives de la qualité vidéo perçue en télévision numérique par câble*.

Recommandation UIT-R BT.500, *Méthodologie d'évaluation subjective de la qualité des images de télévision*.

Recommandation UIT-T J.340 (2010), *Algorithme de référence pour le calcul du rapport signal de crête sur bruit d'une séquence vidéo traitée avec des décalages spatiaux constants et un délai constant*.